

* 2007.02.20 CPB Meeting Followup

See [SamGridListV1](#) for the latest version of the list (dated 2007.02.14). Please refer to this document, as I do not repeat much of its text.

** Preliminary questions

Someone (not me) needs to write requirements documents for the items where none exists. I will make guesses as to the requirements, but someone should write them down separately.

There are many items marked "might need tuning". What does this mean? What are the requirements that tuning is supposed to reach?

*** Skimming in production

There is a skimming requirements document (see <http://www-d0.fnal.gov/computing/reprocessing/primary/>) that I think was written by Daniel last year [Chip, perhaps you can link to this document on Plone]. There is no date on this document - is it still valid? Assuming that it is and that the requirements within still hold, then skimming entails...

Skimming takes merged thumbnails as input.

Skimming runs the skim application to produce many streams of unmerged skimmed thumbnails.

The unmerged skimmed thumbnails may need to be merged before storing to tape (obviously, merging must NOT merge across different streams).

Are these steps correct and still valid?

The document then goes into detail about recovering from failures. Most of the work goes into d0repro tools. Are there any specific tasks for SAMGrid?

What is left to do for SAMGrid here? Time estimate?

What is left to do for d0repro tools? Time estimate?

*** Merge more than one stream from Reco

Is there a requirements document for this task?

The phrase "Tasks 5 and 6 are clearly coupled..." is wrong since the tasks have been renumbered for this document. But indeed for skimming to work, this task must be completed.

What does SAMGrid need to do here? Time estimate?

*** Run raw2sim on minbias files (really raw files from the monitor stream)

A requirements document needs to be written. I think this task involves reading in monitor stream raw data (where here this stream comes from the detector, not reco, and is already stored in SAM with meta-data). The "raw2sim" application is run to create "sim" files, which must be stored back into SAM. Recovery from failures is not critical. Each raw input file produces one sim file. I believe the sim files are big enough that merging will not be required.

So what this really entails is running a different application on raw data and storing output (one stream, no merging, no skimming, no special error recovery).

What development needs to happen for the d0repro tools? Time estimate?

What development needs to happen for SAMGrid? Time Estimate?

I think much of the work has to go into d0repro so that it can run a non-reco application. Then SAMGrid has to handle launching d0repro and storing the output.

*** Production of CAF Trees

We need a requirements document for this task.

MY ASSUMPTIONS: We make CAF Trees from merged skimmed thumbnails. The

output CAF Tree is about the same size as the input merged skimmed thumbnail, so we don't have to merge the CAF Tree Root files.

So this task is very much like the "raw2sim" task above, except that we read in merged skimmed thumbnails, process them with the "CAFTreeMaker" application (or whatever it is called), and then write out CAF Tree root files. The output CAF Tree files can then be stored directly in SAM without further processing.

Here, recovery from failures is critical, since a loss of CAF tree files can represent a loss of luminosity for an analysis. The questions are the same as for the "raw2sim" task above, except that additional work must be done for handling failures.

What is this additional work (for d0repro, for SAMGrid)? Time estimates?

***** Conversion of d0repro tools to use SAM-v7 request system**

We need a requirements document here for this task that should also spell out any additional requirements on the SAMGrid request system.

I think Daniel has to address this one first. In principal, no changes should be necessary for SAMGrid job management itself, except to handle any interface changes to the request system. We need to know very soon if any changes to the request system are necessary.

***** Standalone production of Recocert output**

There is a requirements document (see URL above). Is it still valid?

The recovery from failures is a main issue.

What development needs to happen for the d0repro tools? Time estimate?

What development needs to happen for SAMGrid? Time Estimate?

***** Merging or recocert-files**

We need a requirements document here. In principle, this task is no different from the main merging task, EXCEPT that the merge application is different (someone, not SAMGrid, will have to put together a self-contained executable that can take unmerged recocert files in and write out a merged recocert file) and the data_tiers and file types are different.

What development needs to happen for the d0repro tools? Time estimate?

What development needs to happen for SAMGrid? Time Estimate?

***** MC functionality of v7 d0runjob and SAMGrid**

What additional tests need to be done? What additional functionality may be needed? These items need to be decided now.

***** Ability to write out any stage of output for MC production**

Who is testing this? What are the results?

***** Start MC production at any point with SAM dataset**

I assume this requires some non-trivial changes to d0runjob. What development needs to happen? Time estimate?

I believe the only change necessary for SAMGrid is passing d0runjob the right parameters and validity checking. Is that right? What is the time estimate to do that work?

***** Ability to run trigsim in MC production**

I assume this is really only a change to d0runjob (except that SAMGrid may have to pass more information to d0runjob). What is the time estimate for this change in d0runjob?

What does SAMGrid have to do in order to accommodate the d0runjob changes?

***** Ability to generate CAF trees in MC production**

Again, I think this is mostly a change to d0runjob to add the CAF tree maker.

Would we want to write out both CAF trees and thumbnails? If we have "ability to write out any stage of output..." working, then this should be only a minimal change for SAMGrid.

***** Ability to run recocert in MC production**

I think this task is nearly identical to "Ability to generate CAF trees..."

Note that I assume that making CAF trees and recocert files occur in SERIAL with the MC production (that is that they are steps of the serial workflow, not done in parallel. While the later could be done, that's probably very difficult to manage and I don't think is necessary for MC production).

**** Development Roll-up**

Going through this list, it seems clear that there may be major development work necessary for d0repro tools and d0runjob. Most of the changes to SAMGrid are in response to interface changes in these tools. I think the only really major development necessary for SAMGrid is to handle merging multiple streams. But then again, there will be a LOT of little development necessary to accommodate changes in d0runjob and d0repro tools.

I think it will be difficult to estimate the work and time needed for many of the SAMGrid tasks until we better understand the interface changes in d0repro tools and d0runjob.

**** Followup**

- Can everyone involved please check my assumptions and send corrections? I have enabled commenting and reviewing for this PDF file. If you use Adobe Acrobat (even the free version), you can make comments. See Comments in the Acrobat help. If you are viewing this PDF in your web browser, you will probably have to save it to your disk and load it in Acrobat proper to see the commenting tools. Send the new version back to me. Or just e-mail me text of corrections.
- Can Daniel (for d0repro) tools and Peter Love/Michel Jaffre (for d0runjob) please comment on work to be done and time estimates?
- Can Gabriele/Parag please give work to be done and time estimates for SAMGrid? What tasks do you need more information from d0repro tools/d0runjob before you can make an estimate?